

**A SEMANTIC INFORMATION RETRIEVAL APPROACH TO SOLVING
PAPER-REVIEWER ASSIGNMENT PROBLEM USING A NEURAL NETWORK
LANGUAGE MODEL**

BY

OGUNLEYE, OLAWOLE MOSES

(15PCG01035)

**A DISSERTATION SUBMITTED IN THE DEPARTMENT OF COMPUTER AND
INFORMATION SCIENCES, TO THE SCHOOL OF POSTGRADUATE
STUDIES, COVENANT UNIVERSITY, OTA, OGUN STATE.**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD
OF MASTER OF SCIENCE DEGREE IN COMPUTER SCIENCE.**

JUNE, 2017

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfillment of the requirements for the award of Masters of Science degree in Computer Science in the Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Ogun State.

Mr. Philips John Akinokhai

Secretary, School of Postgraduate Studies

Signature and Date

Prof. Samuel Wara

Dean, School of Postgraduate Studies

Signature and Date

DECLARATION

I hereby declare that Ogunleye, Olawole Moses with matriculation number 15PCG01035, carried out this research entitled “A Semantic Information Retrieval Approach to Solving Paper-Reviewer Assignment Problem Using a Neural Network Language Model”. The project is centered on an original study in the department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, under the supervision of Dr. Adebisi Ayodele. Concepts of this research project are results of the research carried out by Ogunleye, Olawole Moses and ideas of other researchers have been fully recognized.

Signature:-----

Date:-----

CERTIFICATION

This is to certify that this research entitled “**A Semantic Information Retrieval Approach to Solving Paper-Reviewer Assignment Problem Using a Neural Network Language Model**” was carried out by Olawole Moses Ogunleye with matriculation number 15PCG01035 under our supervision and approved by us:

Dr. Ayodele A. Adebiyi
Supervisor

Signature and Date

Dr. Olufunke Oladipupo
Head of Department

Signature and Date

External Examiner

Signature and Date

DEDICATION

I dedicate this project to God Almighty for His sufficient grace, wisdom and knowledge given to me throughout my Master's Degree Programme.

ACKNOWLEDGEMENTS

I would like to use this opportunity to thank the helpers of my destiny whose immense contribution has made this work come to a reality.

First, I want to thank my supervisor, Dr. Adebisi Ayodele, for his insight and encouragement in guiding me through to the completion of this work.

Secondly, I want to thank the Head of Department, Computer and Information Sciences, Dr. Olufunke Oladipupo, for giving me the special privilege to pursue my passion and interest.

Thirdly, I wish to thank the members of faculty, Dr. Mrs. Afolabi, Mr. Stephen Adubi, Ms. Olaperi Sowunmi, who have contributed their insights, expertise and time to making this work a reality.

Also, I will like to thank my colleagues, Sam Ojumah, Anniekeme Udoh, Oluwale Olajide, Aneh Farida, Promise Odiah, Simisola, Tope Ifebanjo, Theresa Abiodun, Deji Osanyi, Dele, Popoola. Not forgetting my colleagues at my work place, Pastor Sola Olayanju, Pastor Oreva Iboje, Pastor Laura O., and Pastor Israel O.

I want to specially thank the Chancellor and Management of Covenant University, for giving me the platform and enabling environment to pursue God's agenda for my life in academics.

I will not fail to mention my loving parents, Pastor and Dr. Mrs. Caleb Olusiji Oluleye, for their continuous support and encouragement that has propelled me to always pursue excellence. I will like to thank my special siblings Engr. David Olabisi Oluleye, Fife Oluleye, Pomile Oluleye and to my special angel, Iretiola Hameed for your benevolent help and holding me strong.

Finally, I thank my heavenly Father and Creator who has led and guided me as I worked through this research.

TABLE OF CONTENTS

Title	Page
Acceptance	ii
Declaration	iii
Certification	iv
Dedication	v
Acknowledgements	vi
Table Of Contents	vii
List Of Figures	x
List Of Tables	xii
Abstract	xiii

CHAPTER ONE: INTRODUCTION

1.1	Background Information	1
1.2	Aim and Objectives	3
1.3	Research Methodology	5
1.4	Significance of the Study	6
1.5	Contribution to Knowledge	6
1.6	Limitation of the Study	6
1.7	Definition of Terms	7
1.8	Organization of the Dissertation	7

CHAPTER TWO: LITERATURE REVIEW

2.1	Introduction	8
2.2	Artificial Intelligence	8
2.2.1	Neural Networks	9
2.2.2	Artificial Neuron	9
2.2.3	Neural Networks Architecture	10
2.2.4	Feed-Forward Neural Network	10
2.2.5	Components Of A Multi-Layer Perceptron	11
2.2.6	How the Back Propagation Algorithm Works	12
2.3	Natural Language Processing	18
2.4	Information Retrieval	19
2.4.1	Information Retrieval Process	20
2.4.2	Document Representations	21
2.5	Information Retrieval Models	25
2.5.1	Exact Matching Models	25

2.5.2	Vector Space Models	25
2.5.3	Probabilistic Models	26
2.5.4	Evaluation Metrics	31
2.6	Paper-Reviewer Assignment	32
2.6.1	Topical Models	33
2.6.2	Optimization Algorithms	33
2.7	Related Works	34

CHAPTER THREE: SYSTEM DESIGN AND METHODOLOGY

3.1	Introduction	36
3.1	Methodology	36
3.1.1	Data Collection And Processing	36
3.2	Similarity Computation	38
3.2.1	Term Frequency-Inverse Document Frequency (TF-IDF)	39
3.2.2	Latent Semantic Indexing (LSI)	40
3.2.3	Latent Dirichlet Allocation (LDA)	41
3.2.4	Word2vec	42
3.3	Assignment Optimization	43
3.4	Overview Of The Proposed System	44
3.5	Paper-Reviewer Assignment System Framework	44
3.6	Requirements Specification	45
3.6.1	Functional Requirements	45
3.6.2	Non-Functional Requirements	46
3.7	Requirements Analysis	46
3.8	The Proposed Paper-Reviewer Assignment System Architecture	49
3.8.1	Use Case Model	51
3.9	Data Modelling	53
3.9.1	Database Design	53
3.9.2	Entity Relationship Modeling	55

CHAPTER FOUR: SYSTEM IMPLEMENTATION, RESULTS AND DISCUSSION

4.1	Implementation Description Processes And Tools	58
4.1	The Implementation Tools Used	58
4.1.1	Programming Language	58
4.1.2	The Web Server	58
4.1.3	The Database Management System	58
4.1.4	Other Libraries	59
4.2	System Development	59

4.2.1	Building The Google Scholar Script	59
4.2.2	Computing the Suitability Scores	60
4.2.3	Building the Assignment Module	61
4.3	Screenshot of The System	62
4.3.1	Login Page	62
4.3.2	Home Page	62
4.3.3	Reviewers Management	63
4.3.4	Create Scoring Job	64
4.3.5	Results Page	65
4.4	Experimentation	66
4.5	Result Findings	77
4.6	Discussion	78
CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATION		
5.1	Summary	79
5.1.1	Contribution to Knowledge	80
5.2	Conclusion	80
5.3	Recommendation	80
	References	81
	Appendix A	86
	Appendix B	111

Table of Figures

Figure	Page
Figure 2.1: A biological neuron	9
Figure 2.2: An Artificial neuron	9
Figure 2.3: Feed-Forward Neural Network	10
Figure 2.4: How the Back Propagation algorithm works	13
Figure 2.5 – Simple recurrent neural network	18
Figure 2.6: The information retrieval process	21
Figure 2.7: The Graphical representation of the LDA model	24
Figure 2-7: architecture of a Word2Vec Framework model	29
Figure 2-8: CBOW model	30
Figure 2-9: Skip-gram model	31
Figure 3.1: Overview of the methodology	36
Figure 3.3: Three-tier Architecture of the System	50
Figure 3.3: The 3-tiered architecture of the PRA System	50
Figure 3.4: Use case diagram for Administrator and the Reviewer	51
Figure 3.5: Use case diagram for Organizers	52
Figure 3.6: Entity relationship diagram for the system database	56
Figure 3.7 Pseudo code of the Paper-Reviewer Assignment System	57
Figure 4.3: Login page	62
Figure 4.4: Home Page	63
Figure 4.6: Compute Scoring Session Page	64
Figure 4.7: Compute Scoring Session processing Page	64
Figure 4.8: The generated scores list	65
Figure 4.9: The generated scatter plot	65
Figure 4.10: The paper-reviewer assignment list	66
Figure 4.11: Depicts the accuracy of the models	67
Figure 4.10: TFIDF Scatter Plot	70
Figure 4.11: LSI Scatter Plot	71
Figure 4.12: LDA Scatter Plot	72
Figure 4.13: Word2Vec Scatter Plot	73
Figure 4.14: Marked line scatter plot of Mean of results	74
Figure 4.15: Marked line scatter plot of Mean of R1	74

Figure 4.16: Marked line scatter plot of Mean of R2	75
Figure 4.18: Marked line scatter plot of Mean of R4	76
Figure 4.19: Marked line scatter plot of Mean of R5	76

LIST OF TABLES

Table	Page
Table 2.1: Different Versions of Back Propagation Neural Networks	15
Table 2.2: Other applications of feed forward neural network	17
Table 3.1: Requirements and Description	47
Table 3.2: Module and Requirements Supported	48
Table 3.3: reviewer user information table	53
Table 3.4: Reviewer Information Table	53
Table 3.5: Reviewer_repo information table	54
Table 3.6: Reviewer_scoring_session information table	54
Table 3.7 Reviewer_scoring_session_results information table	55
Table 4.1: Listing the total average scores for average rankings	67
Table 4.2: Assignment comparison of the models based on-ground truth	68
Table 4.3: The Reviewers Table	69
Table 4.4: The Assignment Optimization Constraints and respective values	69
Table 4.5: Summary of Mean and Standard Deviation of results	77
Table 4.3 Generated suitability Scores based on TF-IDF	86
Table 4.4: Optimized paper-reviewers assignment based on TF-IDF Scores	87
Table 4.5: Generated suitability scores based on LDA Model	91
Table 4.6: Optimized paper-reviewers assignment based on LDA Model Scores	94
Table 4.7: Generated suitability scores based on LSI Scores	99
Table 4.8: Optimized paper-reviewers assignment based on LSI Model Scores	100
Table 4.9: Generated suitability scores based on Word2Vec Scores	105
Table 4.10: Optimized paper-reviewers assignment based on Word2Vec Model Scores	106

ABSTRACT

The task of assigning papers to reviewers is crucial to the realisation of an effective peer review of academic conferences and journal articles. For an excellent performance of the reviewers, it is important that the papers assigned to them are related to their knowledge domain. The manual process of ensuring paper submissions assigned to reviewers is related to the reviewers' knowledge domain can be very cumbersome and inefficient. Besides, low quality and unfair reviews can result from an inefficient assignment of papers to the invited reviewers.

From extant literature, automated reviewer assignment systems have been built to address this challenge as an information retrieval problem. In a bid to leverage on the recent advancement of artificial neural networks in solving natural language problems in the society, a neural network language model, Word2Vec was used to derive suitability scores based on the semantic relatedness between a submitted paper meant for review and a reviewer's representation. The Integer Linear Programming model used suitability scores to optimise the assignments to ensure the workload on the reviewers is balanced.

To test our model and compare with other Information retrieval models used in solving paper-reviewer assignment problem, a system was implemented in Python programming language. Python libraries such as Natural Language Toolkit was used to perform natural language processing on the experimental datasets and the Gensim Python library were used to implement the models. ORTools, a python library for operation research, was used to develop the Integer Linear Programming exact optimisation used for the assignments. Django framework was used to make the project portable for the web. MySQL was the database management system used in managing the database. Celery was used to make the large tasks run on the background.

From our experiment, we explored whether Word2Vec could be used in paper-reviewer assignments and our results indicate that Word2Vec had approximately the same performance with Latent Semantic Indexing. Evaluation was carried out to test the efficiency of the assignment system on an on-ground truth basis. The results show that Word2Vec provided more accuracy in semantic assignments than Latent Semantic Indexing.